SSOAR
Open Access Repository
www.ssoar.info

Multidimensional scaling with regional restrictions for facet theory: an application to Levy's political protest data

Groenen, Patrick J.F.; Lans, Ivo A. van der

Veröffentlichungsversion / Published Version
Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Groenen, P. J., & Lans, I. A. v. d. (2006). Multidimensional scaling with regional restrictions for facet theory: an application to Levy's political protest data. In M. Braun, & P. P. Mohler (Eds.), *Beyond the horizon of measurement: Festschrift in honor of Ingwer Borg* (pp. 41-64). Mannheim: GESIS-ZUMA. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49168-2

Nutzungsbedingungen:
Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.
By using this particular document, you accept the above-stated conditions of use.

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Multidimensional Scaling with Regional Restrictions for Facet Theory: An Application to Levy's Political Protest Data

Patrick J.F. Groenen & Ivo A. van der Lans

Abstract: *Multidimensional scaling (MDS) is often used for the analysis of correlation matrices of items generated by a facet-theory design. The emphasis of the analysis is on regional hypotheses on the location of the items in the MDS solution. An important regional hypothesis is the axial constraint, where the items from different levels of a facet are assumed to be located in different parallel slices. The simplest approach is to do an MDS and draw the parallel lines separating the slices as good as possible by hand. Alternatively, Borg & Shye (1995) proposed to automate the second step. Borg & Groenen (1997, 2005) proposed a simultaneous approach for ordered facets, when the number of MDS dimensions equals the number of facets. In this paper, we propose a new algorithm that estimates an MDS solution subject to axial constraints without the restriction that the number of facets equals the number of dimensions. The algorithm is based on constrained iterative majorization of De Leeuw & Heiser (1980) with special constraints. This algorithm is applied to Levy's (1983) data on political protests.*

Introduction

Multidimensional scaling (MDS) has long been an important technique for analyzing data obtained with facet theory (FT, see, for example, Borg & Shye 1995). Amongst other areas, Ingwer Borg has been advocating these two methods as useful tools for theory building and data analysis. A strong point of FT is the careful design by which items are constructed. Often, correlations between these items are visualized by MDS. Consider Table 1 that shows a dissimilarity matrix **Δ** and a facet design of three facets. In practical facet-theory applications, the dissimilarities are often transformed correlations r_{ij} between items: $\delta_{ij} = 1 - r_{ij}$. The facets can be viewed as categorical design variables on the items of the analysis. The first facet in Table 1 divides the items into three different categories, the second facet into two categories and the third into three. Thus, every item j belongs to a single category on each of the three facets. In facet theory, it is typically postulated that the facets imply particular structures on the empirical intercorrelations, and that these structures will be reflected by regional hypotheses in the MDS solution. Each facet is assumed to partition the MDS space into regions in one of three manners (see Figure 1): by an *axial* partitioning (division into slices along a line), a modular partitioning (division into concentric bands) or a polar partitioning (division into pie pieces), see Guttman (1959), Borg & Shye (1995), and Borg & Groenen (1997, 2005). Note that the lines in Figure 1 are drawn by hand. In this paper, we limit ourselves to axial partitioning only.

Figure 1 Three ways of partitioning the MDS space for a three level facet

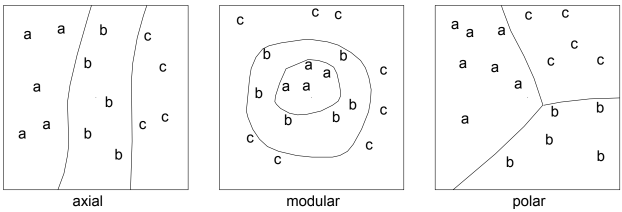


Table 1 Example of a dissimilarity matrix Δ and a three facet design

Dissimilarity matrix **Δ**

	I_1	I_2	I_3	$\cdots$	I_{n-1}	I_n
I_1	0					
I_2	δ_{12}	0				
I_3	δ_{13}	δ_{23}	0			
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$		
I_{n-1}	$\delta_{1,n-1}$	$\delta_{2,n-1}$	$\delta_{3,n-1}$	$\cdots$	0	
I_n	δ_{1n}	δ_{2n}	δ_{3n}	$\cdots$	δ_{2n}	0

Facet design

	Facet 1	2	3
I_1	a	a	c
I_2	a	b	c
I_3	b	a	c
$\vdots$	$\vdots$	$\vdots$	$\vdots$
I_{n-1}	c	b	a
I_n	c	b	a

Given some MDS solution, the construction of the regions is most often done subjectively by the researcher and not through a standardized computational method. A first step towards imposing regions automatically was proposed by Borg & Shye (1995) who provided a two step procedure: first do an MDS followed by finding an optimal location of the lines separating the categories of a facet. In Borg & Groenen (1997, 2005), a simultaneous approach was proposed that searched for an MDS solution that is constrained such that the levels of a facet are linearly separated. The advantage of this approach is that only theory-consistent MDS solutions are considered and that the program Proxscal in SPSS (Meulman, Heiser & SPSS 1999) can handle these constraints. However, the application of the simultaneous approach is limited to situations in which the number of dimensions is the same as the number of facets. Furthermore, Proxscal only works for ordered facets.

In this paper, we propose a Multiple Axial-wise Partitioning Constraints MDS model (MAPC). The MAPC MDS model makes it possible to incorporate axial facet constraints in the construction of the MDS map. An important characteristic of our MAPC MDS model is that it is able to handle situations in which there are more constraining facets than dimensions, so that the dimensionality and therefore the complexity of the interpretation of the MDS map can be kept at a minimum. In addition, MAPC MDS can handle unordered facets as well.

This paper is organized as follows. First, we introduce the main ideas of regional constraints in MDS using axial partitioning. Then, we derive the MAPC MDS algorithm and apply the method to data of Levy on political protests. We end the paper with some conclusions.

Multiple axial-wise partitioning constraints

We propose a multiple axial-wise partitioning constrained MDS for the analysis of data obtained by facet-theory design that is able to deal with more (categorical) facets than dimensions. The basic idea underlying the MAPC model is that we require the MDS solution to be partitioned into successive slices for each facet. Items from different levels of facets are required to lie in different slices, whereas items from the same level are required to lie within the same slice. The hyperplanes separating the slices corresponding to a single facet are parallel to each other, which implies that they are orthogonal to one particular direction in the MDS space.

Table 2 Example of a facet design H

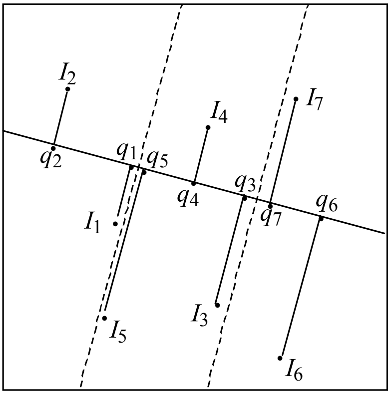
	$\mathbf{h}_1$	$\mathbf{h}_2$	$\mathbf{h}_3$
I_1	1	1	2
I_2	1	2	2
I_3	2	1	3
I_4	2	2	1
I_5	2	1	3
I_6	3	1	3
I_7	3	2	1

Consider the small example of $m = 3$ facets in the $n \times m$ matrix **H** in Table 2. Let us focus on the first facet $\mathbf{h}_1$ which has three levels. Assume for the moment that this facet is ordered, that is, the axial hypothesis implies that the items are located in subsequent parallel regions that are ordered along a line according to the order of the levels. Because the regions are parallel, two adjacent regions representing adjacent levels of the facet $\mathbf{h}_1$ are separated by parallel hyperplanes that form parallel lines in 2D. Orthogonal to these separation hyperplanes, a line exists such that the orthogonal projection of the items onto that line satisfy the set of inequalities

$$q_1, q_2 \leq q_3, q_4, q_5 \leq q_6, q_7. \quad (1)$$

Figure 2 shows an MDS solution that is consistent with facet $\mathbf{h}_1$ treated ordinally. Any MDS solution for which the projections of the items on some line in the MDS space satisfy (1) yields an axial partitioning for the ordinal facet $\mathbf{h}_1$.

Figure 2 **Example of an axial partitioned MDS space where the items of three levels are separated by an axial partitioning. The projections q_i of the items on the line orthogonal to the dashed separation lines satisfy the set of inequalities in (1)**



What if the regions of the facet are not ordered, that is, if the levels of the facet are treated ordinally? Then, any permutation of the labels of $\mathbf{h}_1$ is equally good. Therefore, for a facet $\mathbf{h}_1$ with three levels there are six admissible projections on the line:

$$
\begin{array}{rll}
q_1, q_2 & \leq q_3, q_4, q_5 & \leq q_6, q_7 \\
q_1, q_2 & \leq q_6, q_7 & \leq q_3, q_4, q_5 \\
q_6, q_7 & \leq q_1, q_2 & \leq q_3, q_4, q_5 \\
q_6, q_7 & \leq q_3, q_4, q_5 & \leq q_1, q_2 \\
q_3, q_4, q_5 & \leq q_1, q_2 & \leq q_6, q_7 \\
q_3, q_4, q_5 & \leq q_6, q_7 & \leq q_1, q_2.
\end{array}
$$

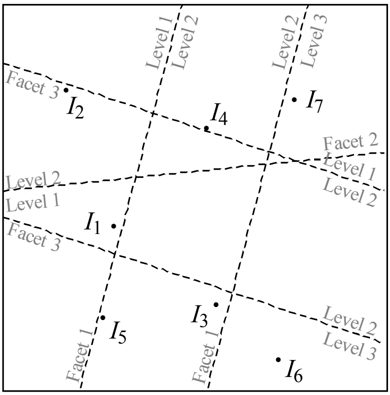
For such a nominal facet, we one can run MDS subject to all six orders and retain the best solution.

In terms of matrix algebra, we can specify multiple axial constraints as follows. Let **X** be the $n \times p$ matrix of coordinates of n items in p dimensions. For multiple facets, let **Q** be the $n \times m$ matrix with the projections for each of the m facets. Then, the coordinates **X** are restricted by

$$\mathbf{X} = \mathbf{QC} \tag{2}$$

subject to rank of **Q** is p and **C** is an $m \times p$ matrix. The combination of the inequality restrictions such as (1) and the rank p restriction on **Q** ensures that the regional constraints of all facets are simultaneously satisfied. Figure 3 shows an example of an MDS solution where the items satisfy the restrictions of the multiple axial regions for the three facets in Table 2.

Figure 3 A multiple axial partitioned MDS space satisfying the axial partitioning restrictions imposed by the three facets in Table 2. The dashed lines are the separation lines



Note that not all combinations of the levels of the three facets are present in Table 1, as there are seven items with different combinations of facet levels out of a possible number of 3×2×3 = 18 different combinations. Yet Figure 3 allows us to reconstruct 14 different regions, of which only 13 are visible in the figure, that correspond to 14 combinations of the three facet levels. The location of a level-separation line is uniquely determined only when projections of at least two items from each of the levels onto the characteristic's vector coincide. If that is not the case, then there is still some freedom left. Instead of level-separation lines, we will have level-separation regions. As the number of levels within a facet increases or the number of facets increase, then there is usually not much freedom left to place the separation line.

The Stress function of MDS can be very flat near the (local) minimum, or can have various similar local minima for completely different configurations (Borg & Lingoes 1980). The idea behind constrained MDS methods in general is that constrained configurations can be interpreted in terms of the external information (constraining attributes), with possibly an only slightly lower fit than the unconstrained configurations. That is, we may get a more easily to interpret configuration without having to offer too much precision in the representation of preferences or dissimilarities. In practice, it is advisable to always compare MAPC-constrained MDS configurations with their unconstrained counterparts.

An algorithm for imposing MAPC in MDS

The multiple axial partitioning constraints do *not* operate on each dimension separately. Instead, coordinates on one dimension affect the feasible set of coordinates on other dimensions. As a consequence of the dependency of dimensions with respect to the constraints, MDS with multiple axial partitioning constraints can not be computed by simple extensions of previously proposed algorithms for other constrained MDS methods, like MDS with dimension-wise order constraints on the coordinates (see Heiser & Meulman 1983; Borg & Groenen 1997) and dimension-wise monotone-spline constraints as proposed by Winsberg & De Soete (1997). Nor can MAPC MDS be implemented by existing algorithms for MDS methods that do impose constraints simultaneously on all dimensions, like MDS with constraints on interpoint distances (Skarabis 1978; Borg & Lingoes 1980), with “circle constraints” (Borg & Lingoes 1980; De Leeuw & Heiser 1980), more general equality and inequality constraints (Lee 1984), configuration-size constraints (Mathar 1990) or reduced-rank subspace constraints for subsets of objects (Borg 1977).

For finding MDS solutions under multiple partitioning constraints, we use the iterative majorization algorithm of De Leeuw & Heiser (1980), called SMACOF. Their iterative majorization algorithm minimizes the following weighted least squares Stress function

$$\sigma^2(\mathbf{X}) = \sum_{j=1}^{n} \sum_{i=j+1}^{n} w_{ij} \, (\delta_{ij} - d_{ij}(\mathbf{X}))^2 \tag{3}$$

under any set of constraints, by finding a series of constraints-satisfying configurations with monotonically decreasing Stress values. In (4), $d_{ij}(\mathbf{X})$ is the Euclidean distance between objects i and j in a p-dimensional space, whose coordinates are given in rows i and j of the $n \times p$-configuration matrix $\mathbf{X}$ and w_{ij} is a fixed (nonnegative) weight that weights the contribution of the squared residual of object pair (i,j) to the overall Stress. Instead of $\sigma^2(\mathbf{X})$, we report $\sigma_n^2(\mathbf{X}) = \sigma^2(\mathbf{X}) \big/ \sum_{j=1}^{n} \sum_{i=j+1}^{n} w_{ij}\delta_{ij}^2$, which has the same local minima and the advantage that at a local minimum $\sigma_n^2(\mathbf{X})$ values are always between 0 and 1 and are equal to the square of Kruskal's (1964) Stress-1 (see, Borg & Groenen 2005: 249–250).

In case of MDS for the analysis of intercorrelations between items from a facet design, the n objects correspond to the n items and all w_{ij}'s are larger than 0 (usually equal to 1), unless the intercorrelations for some pairs (i,j) are missing, in which case their associated weights become 0. The summation in (3) is across all lower-diagonal elements of the matrix of dissimilarities $\mathbf{\Delta}$ only, because this matrix is assumed to be symmetric.

In each step of the iterative majorization algorithm, a better constrained configuration is found by constructing a function $\varphi(\mathbf{X},\mathbf{X}^*)$ that majorizes $\sigma^2(\mathbf{X})$, and by minimizing $\varphi(\mathbf{X},\mathbf{X}^*)$ over $\mathbf{X}$. A function $\varphi(\mathbf{X},\mathbf{X}^*)$ is said to majorize function $\sigma^2(\mathbf{X})$, when $\sigma^2(\mathbf{X}) \leq \varphi(\mathbf{X},\mathbf{X}^*)$ for all $\mathbf{X}$, and $\varphi(\mathbf{X},\mathbf{X}^*) = \sigma^2(\mathbf{X})$ when $\mathbf{X} = \mathbf{X}^*$. Let $\hat{\mathbf{X}}$ be the constrained $\mathbf{X}$ that minimizes $\varphi(\mathbf{X},\mathbf{X}^*)$ for some $\mathbf{X}^*$, then we have the following set of (in)equalities:

$$\sigma^2(\hat{\mathbf{X}}) \leq \varphi(\hat{\mathbf{X}},\mathbf{X}^*) \leq \varphi(\mathbf{X}^*,\mathbf{X}^*) = \sigma^2(\mathbf{X}^*)$$

Clearly, choosing $\mathbf{X}^*$ to be equal to the constrained configuration obtained in the previous iteration and minimizing $\varphi(\mathbf{X},\mathbf{X}^*)$ over $\mathbf{X}$ subject to the constraints, we obtain a constrained configuration with lower (or equal) Stress in each iteration. $\mathbf{X}^*$ is called a supporting point of the majorizing function. Because the series of Stress values of successive constrained configurations is monotonically decreasing and the Stress is bounded from below by zero guarantees that the algorithm will converge to at least a local minimum. For more information on iterative majorization, see, for example, De Leeuw (1994), Heiser (1995) or, for an introduction, Borg & Groenen (2005).

Majorization is useful when a complicated function, like $\sigma^2(\mathbf{X})$ can be majorized by a simpler function. A relatively quadratic function in $\mathbf{X}$ that majorizes $\sigma^2(\mathbf{X})$ is

$$\begin{aligned}\varphi(\mathbf{X},\mathbf{X}^*) &= \sum_{j=1}^{n}\sum_{i=j+1}^{n}\left(w_{ij}\delta_{ij}^2 - 2b_{ij}(\mathbf{X}^*)(\mathbf{x}_i - \mathbf{x}_j) + w_{ij}d_{ij}^2(\mathbf{X})\right) \\ &= c_1 + \|\mathbf{X} - \mathbf{V}^{-}\mathbf{B}(\mathbf{X}^*)\mathbf{X}^*\|^2_{\mathbf{V}},\end{aligned}$$

where

$\mathbf{x}_i$	is the p-vector containing elements from the i^{th} row of $\mathbf{X}$,
$b_{ij}(\mathbf{X}^*)$	$= w_{ij}\delta_{ij}/d_{ij}(\mathbf{X}^*)$, if $d_{ij}(\mathbf{X}^*) > 0$, and
$b_{ij}(\mathbf{X}^*)$	$= 0$, if $d_{ij}(\mathbf{X}^*) = 0$,
$\|\mathbf{Z}\|^2_{\mathbf{V}}$	denotes the weighted squared Euclidean norm given by tr $(\mathbf{Z}'\mathbf{V}\mathbf{Z})$,
$\mathbf{V}$	is an $n\times n$-matrix, whose elements v_{ij} are equal to $-(w_{ij}+w_{ji})$ if $i \neq j$, and whose elements v_{ii} are equal to $\sum_{j\neq i} w_{ij}$,
$\mathbf{V}^{-}$	is a generalized inverse of $\mathbf{V}$,
$\mathbf{B}(\mathbf{X}^*)$	is the $n \times n$ matrix with elements $-b_{ij}$ if $i \neq j$ and $\sum_{j\neq i} b_{ij}$ if $i = j$,
c_1	is a constant that depends on $\mathbf{X}^*$.

The unconstrained $\mathbf{X}$ that minimizes $\varphi(\mathbf{X},\mathbf{X}^*)$ which we will refer to by $\overline{\mathbf{X}}$, that is equal to $\mathbf{V}^{-}\mathbf{B}(\mathbf{X}^*)\mathbf{X}^*$. The constrained $\mathbf{X}$ that minimizes $\varphi(\mathbf{X},\mathbf{X}^*)$ can be found by projecting $\overline{\mathbf{X}}$ onto the multiple axial constraints in the metric defined by $\mathbf{V}$.

In the previous section, we stated that the multiple axial partitioning constraints can be imposed by requiring $\mathbf{X} = \mathbf{QC}$ such that $\mathbf{q}_k$ satisfies the partitioning constraints of each facet k and $\mathbf{Q}$ being of rank p. To avoid the possibility that some $\mathbf{q}_k = \mathbf{0}$, we also impose additionally the constraint $\|\mathbf{q}_k\|^2 = n$. These restrictions are equivalent to $\mathbf{X} = \mathbf{ZA'C}$ subject to the constraints $\mathbf{Za}_k \in C_k$ and $\|\mathbf{Za}_k\|^2 = n$. Here,

$\mathbf{Z}$	is an $n \times p$-matrix of to be estimated parameters,
$\mathbf{A}$	is an $m \times p$-matrix of to be estimated parameters,
$\mathbf{a}_k$	is the m-vector containing the elements from the kth row of $\mathbf{A}$,
$\mathbf{C}$	is an $m \times p$-matrix of to be estimated parameters, and
C_k	is the closed convex cone that denotes the inequality constraints that are implied by the kth level.

So, the matrix $\mathbf{A}$ defines directions in the p-dimensional space given by $\mathbf{Z}$. Writing $\mathbf{X}$ in terms of $\mathbf{ZA'C}$ and imposing constraints on the columns of $\mathbf{ZA'}$ guarantees that the directions given by $\mathbf{C'}(\mathbf{CC'})^{-}$ in the p-dimensional space given by $\mathbf{X}$ satisfy the constraints. The constraints $\|\mathbf{Za}_k\| = n$ are necessary to guarantee that the multiple axial constraints are all satisfied. Without this explicit length constraint a solution can be found that satisfies only p axial constraints with nonzero $\mathbf{a}_k$, and the remaining $m-p$ axial constraints being satisfied by choosing $\mathbf{a}_k = \mathbf{0}$. The inequality constraints defined by each C_k are constraints similar to (1) that typically correspond to a nominal or ordinal measurement level and the primary approach to ties (Gifi 1990; Kruskal 1964; Young 1981). Note that $\mathbf{ZA'}$ is identified up to simultaneous linear transformations of both $\mathbf{Z}$ and $\mathbf{A}$, that is, the combination $\mathbf{Z}$ and $\mathbf{A}$ is equivalent to the combination $\mathbf{ZL}$ and $\mathbf{A}(\mathbf{LL'})^{-1}\mathbf{L}$ for any nonsingular $p \times p$-matrix $\mathbf{L}$. Therefore, without loss of generality, we impose $\mathbf{Z'Z} = n\mathbf{I}$ and $\mathbf{a}_k'\mathbf{a}_k = 1$. Thus, minimizing $\varphi(\mathbf{X},\mathbf{X}^*)$ over the regional restrictions is equivalent to minimizing

$$\varphi_1(\mathbf{Z},\mathbf{A},\mathbf{C}) = \|\overline{\mathbf{X}} - \mathbf{ZA'C}\|^2_{\mathbf{V}}$$

over $\mathbf{Z}$, $\mathbf{A}$, and $\mathbf{C}$, subject to the constraints $\mathbf{Za}_k \in C_k$ and $\|\mathbf{Za}_k\|^2 = n$.

To minimize function $\varphi_1(\mathbf{Z},\mathbf{A},\mathbf{C})$ we use alternating least squares: that is, first we update $\mathbf{C}$, keeping $\mathbf{A}$ and $\mathbf{Z}$ fixed, then we update $\mathbf{A}$, keeping $\mathbf{Z}$ and $\mathbf{C}$ fixed, and then we update $\mathbf{Z}$, keeping $\mathbf{A}$ and $\mathbf{C}$ fixed. These three steps are carried out iteratively. Minimizing $\varphi_1(\mathbf{Z},\mathbf{A},\mathbf{C})$

over **C** is a straightforward regression problem, for which $\hat{\mathbf{C}} = (\mathbf{AZ'VZA'})^{+}\mathbf{AZ'V}\overline{\mathbf{X}}$ is an optimal solution, where $(\mathbf{AZ'VZA'})^{+}$ is the Moore-Penrose inverse, since **AZ'VZA'** is not of full rank. The main difficulty lies in minimizing $\varphi_1(\mathbf{Z},\mathbf{A},\mathbf{C})$ over **Z** and **A** for fixed **C** over the constraint sets C_k and $\mathbf{a}_k'\mathbf{Z'Za}_k = n$.

While updating **Z** and **A**, we keep **C** fixed. Let the singular value decomposition of **C** be given by $\mathbf{P\phi Q'}$, so that the Moore-Penrose inverse of **C** is given by $\mathbf{C}^{+} = \mathbf{Q\phi}^{-1}\mathbf{P'}$. $\varphi_1(\mathbf{Z},\mathbf{A},\mathbf{C})$ can then be written as

$$\varphi_1(\mathbf{Z},\mathbf{A},\mathbf{C}) = \|\overline{\mathbf{X}} - \mathbf{ZA'C}\|^2_{\mathbf{V}} = \|\mathbf{C}^{+\prime}\overline{\mathbf{X}}'\mathbf{V}^{1/2} - \mathbf{AZ'V}^{1/2}\|^2_{\mathbf{CC'}} \tag{4}$$

The metric **CC'** makes it difficult to impose the constraints. Therefore, we use an additional majorization step (e.g. see De Leeuw 1994; Heiser 1995). Let $\mathbf{Y}^*$ be the matrix **ZA'** obtained in the previous iteration, satisfying all constraints. We can then derive a function majorizing $\varphi_1(\mathbf{Z},\mathbf{A},\mathbf{C})$ by using the following inequality:

$$\left\|\mathbf{Y}^{*\prime}\mathbf{V}^{1/2} - \mathbf{AZ'V}^{1/2}\right\|^2_{\mathbf{CC'}-\phi_1^2\mathbf{I}} \leq 0$$

which can be written as

$$\|\mathbf{AZ'V}^{1/2}\|^2_{\mathbf{CC'}} \leq \phi_1^2\,\|\mathbf{ZA'}\|^2_{\mathbf{V}} - 2\mathrm{tr}\,\mathbf{AZ'VY}^{*\prime}(\phi_1^2\,\mathbf{I} - \mathbf{CC'}) + \|\,\mathbf{Y}^{*\prime}\mathbf{V}^{1/2}\,\|^2_{\phi_1^2\mathbf{I}-\mathbf{CC'}}\,, \tag{5}$$

where ϕ_1^2 is the largest eigenvalue of **CC'**, that is, the square of the first diagonal element of $\boldsymbol{\phi}$. Combining (4) and (5) gives

$$\begin{aligned}\varphi_1(\mathbf{Z},\mathbf{A},\mathbf{C}) \leq \phi_1^2\,\|\mathbf{ZA'}\|^2_{\mathbf{V}} &- 2\mathrm{tr}\,\mathbf{AZ'VY}^{*\prime}(\phi_1^2\,\mathbf{I} - \mathbf{CC'}) - 2\,\mathrm{tr}\,\mathbf{AZ'V}\overline{\mathbf{X}}\mathbf{C'} + \\ &\|\mathbf{Y}^{*\prime}\mathbf{V}^{1/2}\|^2_{\phi_1^2\mathbf{I}-\mathbf{CC'}} + \|\overline{\mathbf{X}}\|^2_{\mathbf{V}}.\end{aligned} \tag{6}$$

To simplify notation, let c_2 denote the last two constant terms in (6), and let **M** denote $\mathbf{Y}^*(\mathbf{I} - \phi_1^{-2}\mathbf{CC}) + \phi_1^{-2}\,\overline{\mathbf{X}}\mathbf{C'}$. Then (6) can be written as

$$\varphi_1(\mathbf{Z},\mathbf{A},\mathbf{C}) \leq \phi_1^2 \,||\mathbf{ZA'} - \mathbf{M}||^2_{\mathbf{V}} - \phi_1^2 \,||\mathbf{M}||^2_{\mathbf{V}} + c_2. \tag{7}$$

The right hand side of (7) is a quadratic function in **ZA'**. In the case in which all w_{ij}'s are equal to each other, as is typical in MDS of dissimilarities, **V** can be written as n**J'J**, with **J** (the centering operator) defined as $\mathbf{I} - n^{-1}\mathbf{11'}$. **I** the $n \times n$ identity matrix, and **1** the n-vector with all elements equal to one. Therefore, the first term of the right hand side of (7) can, provided that **Z** and **M** have zero column means which can be imposed without loss of generality, be written as $n\phi_1^2 \,||\mathbf{ZA'} - \mathbf{M}||^2$, which simplifies the majorizing function even further.

For a description of the updates of matrices **A** and **Z,** the reader is referred to the appendix. There are several convergence criteria to be specified for the different loops in the algorithm. First of all, convergence criteria have to be specified for the outer iterations where a constrained configuration with a better Stress value is obtained. Second, a convergence criterion has to chosen for the iterations that minimize $n\phi_1^2 \,||\mathbf{ZA'} - \mathbf{M}||^2$ over **Z** and **A**. Finally, a convergence criterion has to be chosen for the iterations of Dijkstra's cyclic projection algorithm that is used for obtaining the updates of **A** and **Z** described in the appendix.

Regionally constrained MDS for Levy's political protest acts

Levy (1983) studied the attitudes of respondents from different countries towards different protest behaviors. The items Levy considered varied on three facets: (1) the modality of the attitude (evaluation, approval or likelihood of own overt action), (2) the strength of execution (demanding, obstructive or physically damaging), and (3) the way to carry out the protest (omission or commission) and were constructed as follows. First, 10 protest acts were formulated based on combinations of the levels of facets 2 and 3. These 10 protest acts were then combined with all levels of facet 1. This procedure yields the 30 items in Table 3. These items were rated by respondents on a Likert scale ranging from very positive to very negative. The data that we have available are correlation matrices of the items gathered in five countries (1973-1974): Great Britain (n = 1482), Austria (n = 1584), West Germany (n = 2307), The Netherlands (n = 1201), and the United States (n = 1719).

Borg & Groenen (1997, 2005) analyzed these data by ordinal MDS followed by eyeing to trace back the facet structure. Here, we apply metric MACP MDS with nominal facets. Table 4 shows the unconstrained and the MAPC Stress values. For all five countries, there is some extra Stress due to the multiple axial constraints, but not too much. Therefore, we conclude that imposing the regional constraints still fits the data well.

The MAPC solutions for the countries are given in Figures 4 to 6. The maps for Great Britain, the Netherlands, and the United States are pretty much the same in the sense that the levels for Modality and Strength appear in the same order in the MDS solutions. Therefore, only the map for the Netherlands is shown here. It can be seen that items at the Doing level of the Modality facet correlate more with items at the Approve level than with items at the Effective level. Furthermore, items at the Physically damaging level of the Strength facet correlate higher with those at the Demanding level than with those at the Destructive level. It can also be seen that correlations between some protest acts follow more or less the same pattern at different levels of Modality. For instance, Damage and Violence are more correlated with each other than with Slogans on walls, regardless of the level of Modality. For other groups of protest acts, like Block traffic, Lawful demonstrations, and Petitions, this does not hold. It can be seen that the Carrying out facet only makes a further distinction among the protest acts at the Destructive level of the Strength facet. This is a direct consequence of the facet design in which all protest acts at the Physically damaging and the Demanding levels are acts of commission (logically). On closer look, one can see that the line that separates the Omission from the Commission protest acts is not unique. As a matter of fact, the line can be slightly rotated (counter-clockwise) without violating the constraints that are implied by the Carrying out facet.

Table 3 Facet design

Item		Facet A: Modality	Facet B: Strength	Facet C: Carrying Out
1.	Petitions	1 Approve	1 Demanding	2 Commission
2.	Boycotts	1 Approve	2 Obstructive	1 Omission
3.	Lawful demonstrations	1 Approve	1 Demanding	2 Commission
4.	Refusing rent	1 Approve	2 Obstructive	1 Omission
5.	Wildcat strikes	1 Approve	2 Obstructive	1 Omission
6.	Slogans on walls	1 Approve	3 Physically damaging	2 Commission
7.	Occ. buildings	1 Approve	2 Obstructive	2 Commission
8.	Block traffic	1 Approve	2 Obstructive	2 Commission
9.	Damage	1 Approve	3 Physically damaging	2 Commission
10.	Violence	1 Approve	3 Physically damaging	2 Commission
11.	Petitions	2 Effective	1 Demanding	2 Commission
12.	Boycotts	2 Effective	2 Obstructive	1 Omission
13.	Lawful demonstrations	2 Effective	1 Demanding	2 Commission
14.	Refusing rent	2 Effective	2 Obstructive	1 Omission
15.	Wildcat strikes	2 Effective	2 Obstructive	1 Omission
16.	Slogans on walls	2 Effective	3 Physically damaging	2 Commission
17.	Occ. buildings	2 Effective	2 Obstructive	2 Commission
18.	Block traffic	2 Effective	2 Obstructive	2 Commission
19.	Damage	2 Effective	3 Physically damaging	2 Commission
20.	Violence	2 Effective	3 Physically damaging	2 Commission
21.	Petitions	3 Doing	1 Demanding	2 Commission
22.	Boycotts	3 Doing	2 Obstructive	1 Omission
23.	Lawful demonstrations	3 Doing	1 Demanding	2 Commission
24.	Refusing rent	3 Doing	2 Obstructive	1 Omission
25.	Wildcat strikes	3 Doing	2 Obstructive	1 Omission
26.	Slogans on walls	3 Doing	3 Physically damaging	2 Commission
27.	Occ. buildings	3 Doing	2 Obstructive	2 Commission
28.	Block traffic	3 Doing	2 Obstructive	2 Commission
29.	Damage	3 Doing	3 Physically damaging	2 Commission
30.	Violence	3 Doing	3 Physically damaging	2 Commission

Table 4 Stress values per country of unconstrained metric MDS and regionally constrained MAPC MDS with p = 2

Country	Unconstrained Stress	Regionally Constrained Stress
Great Britain	.086	.107
Austria	.084	.107
West Germany	.078	.106
The Netherlands	.087	.104
United States	.076	.107

Figure 4 MAPC MDS solution of the Levy data for Austria

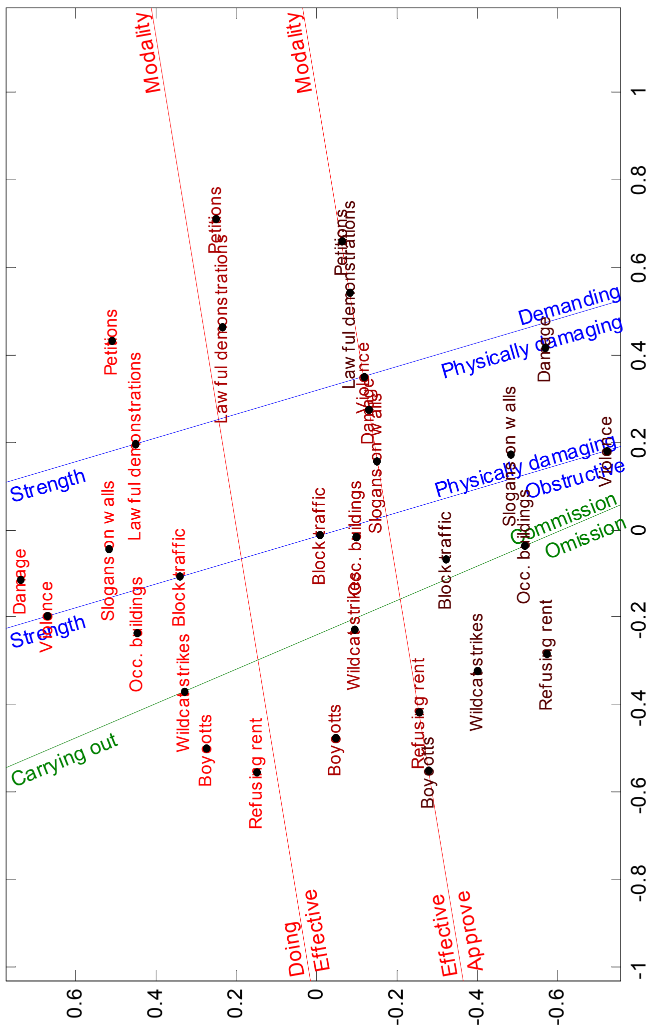


Figure 5 MAPC MDS solution of the Levy data for West Germany

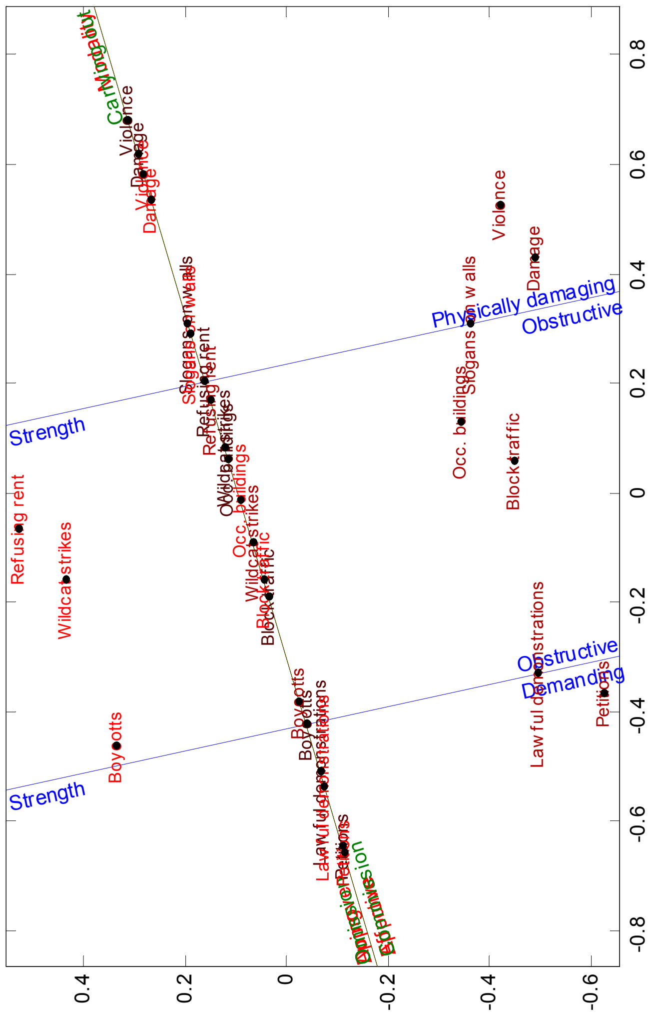
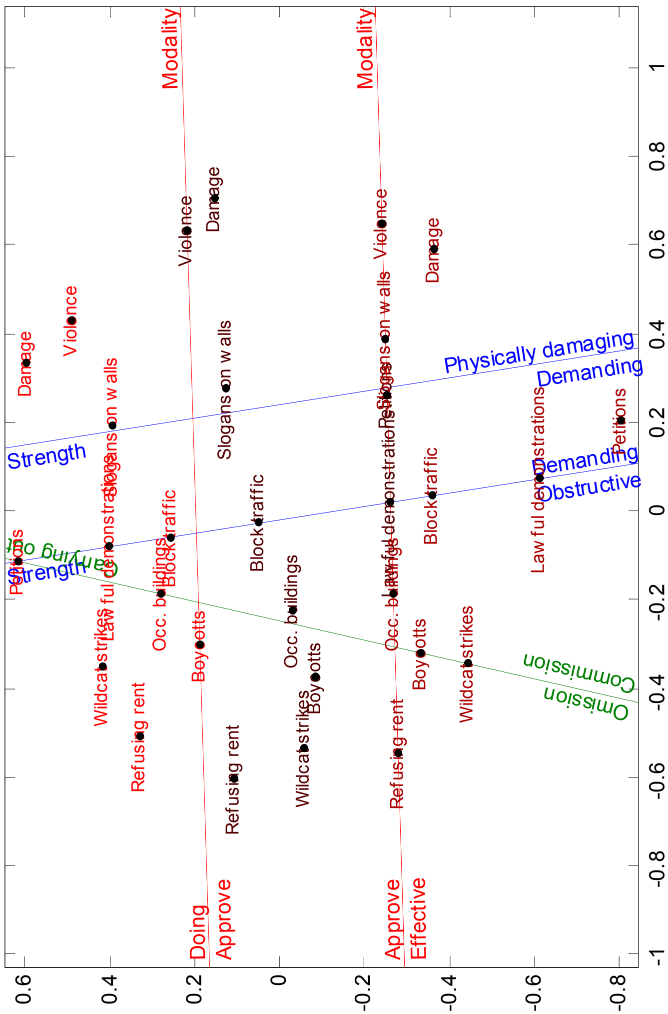


Figure 6 **MAPC MDS solution of the Levy data for The Netherlands**



In the map for Austria, we see an interchange in the order of the Approve and the Effective levels of the Modality facet, and an interchange in the order of the Physically damaging and Demanding levels of the Strength facet. Despite these interchanges, whose interpretation is outside the scope of this paper, the facet structure is very clear.

West Germany shows how things can go wrong. Although the Stress did not increase too much by imposing multiple axial partitioning constraints, it turns out that in the final MDS solution a large number of the constraints imposed by the Modality and the Carrying out facets are active. As such, many items are located on the boundary of their corresponding slices. In fact, the constraints even lead to a situation in which there are no locations in the map that uniquely correspond to the Doing level of the Modality facet. The whole slice coincides with the boundaries of both the Approve and the Effective level. This leads to such a clutter of items in the map that the map becomes virtually uninterpretable. One might argue that it appears that the facet structure apparently is not so dominant in the data. The fact that the Stress increased only slightly seems to be an indication of the flatness of the Stress function near the (perhaps locally) optimal solution. Unlike the other countries, Germany has the Doing and the Obstructive levels of the corresponding facets as middlemost levels.

Conclusion and Discussion

We proposed the MAPC MDS model that allows for the inclusion of multiple axial regional constraints. In MAPC MDS it is possible to specify more constraining facets than dimensions. The facets are incorporated in the MDS map in such a way that each individual level can readily be identified, and there is a unique one-to-one mapping between locations in the map and levels of each of the facets.

The proposed algorithm for MAPC MDS is very flexible in the sense that the Stress function that is being minimized incorporates weights that can be used to control the impact of the misfit of the individual dissimilarities on the overall Stress. Amongst others this makes sure that the algorithm can both be used for MDS of dissimilarities and for the unfolding of preferences. The latter may have strong applications in the area of marketing and new product development.

It must be recognized that the expand-and-shrink operation, rewriting **X** as **ZA'C**, does not completely cover all possible solutions that satisfy the multiple axial-wise partitioning constraints. Actually, it may happen that it is advantageous for the minimization of Stress to confine the constraining directions to a lower-than-p-dimensional subspace of the map. For instance, the advantage of confining the directions to a (p–1)-dimensional subspace will be that the remaining dimension will be completely unrestricted. Actually, our definition of the constraints does not prohibit us to let all directions coincide, to let all items share the same coordinate on that dimension, and to actually fit the dissimilarities in (p–1) unconstrained dimensions! Obviously, such a solution would destroy the whole rationale behind the imposition of multiple axial-wise partitioning constraints. As a matter of fact, the attractiveness (from a Stress minimization point of view) of such uninformative solutions in the illustration shown can be checked by comparing Stress values for unconstrained and MAPC MDS for different dimensionalities.

A two-dimensional perceptual map gives a nice trade-off between an optimal representation of the data and parsimony of the map. The choice of two dimensions made it simple to visualize the level-separating hyperplanes, which are lines in this case, and inspect the MDS maps. Visualization of the level-separating hyperplanes and visual inspection of the maps become much more difficult in higher dimensionality. So, one line of further research could investigate ways in which more-than-two-dimensional perceptual maps with multiple axial partitioning constraints can best be visualized (cf. Buja & Swayne 2002). Preferably one would carry out some empirical studies to see which ways of visualization are most easy to use and most appreciated by the people that will have to deal with the results of MAPC MDS in practice. Of course, results may amongst others depend on the particular type of product and the kind of constraining characteristics involved.

A disadvantage of our multiple axial constraints, dividing the MDS map into slices, is that not every possible level combination of the constraining attributes can be represented by a separate region. Two facets with two levels each yield a region for every combination of levels in two dimensions. However, three facets of two levels each give three separation lines yielding at most seven regions out of eight possible combinations of levels. As the number of facets increase with respect to the number of dimensions, or the number of levels per facet increase, there will be a larger proportion of level combinations that can not be represented by separate regions. As a consequence, more and more level combinations will coincide at level-separation lines and more and more items will be located on these boundaries, making the interpretation increasingly more difficult and the map less useful. Therefore, another line of research may aim at the specification of a different type of constraints that divide the map into mutually exclusive regions in such a way that all (or almost all) level combinations can be represented by a separate region. To keep the interpretation of the map easy, each category combination should preferably correspond to one region and not to two or more disconnected regions. Borrowing ideas from facet theory again, one could think amongst others of radial and polar constraints (e.g. see Borg & Groenen 1997, Chapter 4).

In addition, varieties of MAPC for three-way MDS, such as the weighted Euclidean model (also used in INDSCAL) could be developed. The multiple axial constraints can also be readily implemented in other techniques, like PCA, generalized canonical correlation analysis, correspondence analysis, etc.

We need some procedure that exactly determines whether we have one unique hyperplane or an infinite number of hyperplanes within some ‘hyper region’, separating the subsequent regions that correspond to the levels of a particular facet. To create unique hyperplanes, an extra penalty term could be used much in the same way as in support-vector machines.

Appendix

Updating A

There are several ways to update $\mathbf{A}$. First, we can bring the problem back to a standard nonnegative least squares problem. Minimizing (7) over $\mathbf{A}$ is equivalent to minimizing

$$\tau(\mathbf{a}_k) = \|\mathbf{Z}\mathbf{a}_k - \mathbf{m}_k\|^2, \tag{A1}$$

for each row, $\mathbf{a}_k$, of $\mathbf{A}$. For orthonormal $\mathbf{Z}$, the normalization constraints $\|\mathbf{a}_k\|^2 = 1$ and $\|\mathbf{Z}\mathbf{a}_k\|^2 = n$ are equivalent. Because nominal and ordinal partitioning characteristics define cones, $\tau(\mathbf{a}_k)$ may be minimized without the length constraint followed by proper normalization (De Leeuw 1977; Gifi 1990).

The minization of (A1) can be transformed into a nonnegative least-squares problem as follows. The ordinal partioning constraints are given by $\mathbf{t}_k = \mathbf{S}_k\mathbf{Z}\mathbf{a}_k \geq \mathbf{0}$, so that $\mathbf{S}_k^{-1}\mathbf{t}_k = \mathbf{Z}\mathbf{a}_k$. Thus, minimizing $\tau(\mathbf{a}_k)$ is equivalent to minimizing

$$\| \mathbf{S}_k^{-1}\mathbf{t}_k - \mathbf{m}_k \|^2 \text{ subject to } \mathbf{t}_k \geq \mathbf{0}, \tag{A2}$$

a standard nonnegative least squares problem that can be solved by the analytic method of Lawson & Hanson (1974) or the iterative meтthod of Groenen, Van Os & Meulman (2000).

An alternative approach to updating $\mathbf{A}$ is by using Dykstra's (1983) cyclic projection algorithm. By iterative projections onto the hyperplane defined by $\mathbf{Z}$ and onto the cones C_k, and subsequent proper normalization one obtains the normalized projection, say, $\hat{\mathbf{M}}$, of $\mathbf{M}$ onto the intersection of the hyperplane and the cones. The update for $\mathbf{A}$ is then obtained as $n^{-1}\hat{\mathbf{M}}'\mathbf{Z}$, as $\mathbf{Z}$ is orthonormal. For the projections onto the hyperplane defined by $\mathbf{Z}$, no increment has to be subtracted before projections (see Gaffke & Mathar 1989; von Neumann 1950). The advantage of using the cyclic projection algorithm lies in the fact that projections onto the cones C_k can be carried out very efficiently by Kruskal's (1964) up-and-down blocks algorithm, whereas on the other hand the $\mathbf{S}_k$'s in (A2) will tend to have a large number of rows due to the primary approach to ties. A disadvantage of the cyclic projection algorithm as compared to the analytic method of Lawson & Hanson (1974) is that it is iterative.

Updating $\mathbf{Z}$

To update $\mathbf{Z}$ we first drop the orthonormality constraints on $\mathbf{Z}$, and impose them again afterwards. This procedure works fine, as long as the update without orthonormality constraints is of full rank. We do not expect $\mathbf{Z}$ to be of reduced rank, because that would be contradictory to minimization of the loss unless the dissimilarities can indeed be fitted in a p-dimensional space with $p < m$.

There are again two ways to find an update. First, the problem can be transformed into a standard nonnegative least squares problem. Let $\mathbf{z} = \text{vec}(\mathbf{Z})$ denote the vector of all columns of $\mathbf{Z}$ stacked under each other, and $\mathbf{m} = \text{vec}(\mathbf{M})$ denote the vector of all columns of $\mathbf{Z}$ stacked under each other. Also, let $\mathbf{S}$ denote the partitioned block matrix with blocks equal to $a_{kt}\mathbf{S}_{\text{k}}$, and $\mathbf{G}$ a partitioned block matrix with blocks $a_{kt}\mathbf{I}$, with $\mathbf{I}$ the $n \times n$ identity matrix. Then, minimization of $||\mathbf{ZA}' - \mathbf{M}||^2$ over $\mathbf{Z}$ subject to $\mathbf{S}_k\mathbf{Za}_k \geq \mathbf{0}$ can be written as

$$||\mathbf{Gz} - \mathbf{m}||2 \text{ subject to } \mathbf{Sz} \geq 0.$$

Substituting $\mathbf{t} = \mathbf{Sz}$ gives the standard nonnegative least-square problem of minimizing

$$||\mathbf{GS}^{-1}\text{t} - \mathbf{m}||2 \text{ subject to t} \geq 0.$$

Alternatively Dykstra's (1983) cyclic projection algorithm can be used again. This time, $\mathbf{M}$ is projected onto the intersection of the hyperplane defined by $\mathbf{A}$ and the cones C_k.

Assuming that the solution $\hat{\mathbf{M}}$ is of full rank with singular value decomposition $\mathbf{K\Lambda\tilde{L}'}$ Then choosing

$$\hat{\mathbf{Z}} = n^{1/2}\text{K}$$
$$\hat{\mathbf{A}} = \tilde{\mathbf{A}}\text{L, and}$$
$$\hat{\mathbf{C}} = n^{-1/2}(\tilde{\mathbf{A}}')^{-1}\text{LK}'\hat{\mathbf{M}}\,\tilde{\mathbf{A}}'\tilde{\mathbf{C}}$$

yields an orthonormal update for **Z**, while keeping $\|\mathbf{a}_k\|^2 = 1$. Here, $\hat{\mathbf{Z}}$, $\hat{\mathbf{A}}$, and $\hat{\mathbf{C}}$ are the updates for matrices **Z**, **A**, and **C**, and $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{C}}$ are the matrices **A** and **C** that were kept fixed while updating **Z**.

For nominal axial-wise partitioning constraints we have to check all $\frac{1}{2}\times\Pi_{k=1}^{m} J_k$ combinations of orders along the partition axes, where J_k is the number of levels for facet k. If the J_k or K gets larger, the number of combinations to be checked explodes, so that the algorithm becomes slow. When Dykstra's algorithm is used, a branch and bound strategy can be applied to speed up computations. This branch and bound strategy is based on the fact that, for instance when updating **A**, Dykstra's algorithm yields a series of vectors that are: *i*) successively getting closer to the projection of $\mathbf{m}_k$ onto the intersection of the closed convex cones C_k and the space spanned by **Z**, and *ii*) decreasing in length, whereas this length is a direct measure of the Euclidean norm of the difference between the vector and $\mathbf{m}_k$. So, for any permutation the iterations of Dykstra's algorithm can be stopped, as soon as the length of the vector becomes smaller than the length of the projection obtained for some other permutation. As the overall program approaches convergence it becomes ever more likely that the 'best' permutation from the previous iteration updating **A** will also be the 'best' permutation in the current iteration. Therefore, it is efficient to start with the 'best' permutation from the previous iteration, so that the change of being able to stop Dykstra's algorithm for all other permutations after a relatively small number of iterations is relatively high. While updating **A**, we can apply the branch and bound strategy for the orders from each characteristic, so for each row of **A**, separately. While updating **Z**, we have to apply the strategy for all combinations of orders simultaneously.

References

Borg, I. (1977). Geometric representations of individual differences. In J. C. Lingoes (Ed.), *Geometric Representations of Relational Data* (pp. 609-655). Ann Arbor: Mathesis Press.

Borg, I., & Groenen, P. J. F. (1997). *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer.

Borg, I., & Groenen, P. J. F. (2005). *Modern Multidimensional Scaling: Theory and Applications* (2nd ed.). New York: Springer.

Borg, I., & Lingoes, J. C. (1980). A model and algorithm for multidimensional scaling with external constraints on the distances. *Psychometrika*, *45,* 25-38.

Borg, I., & Shye, S. (1995). *Facet theory: Form and content*. Newbury Park, CA: Sage.

Buja, A., & Swayne, D. F. (2002). Visualization methodology for multidimensional scaling. *Journal of Classification*, *19*, 7-43.

De Leeuw, J. (1977). *Normalized Cone Regression* (Tech. Rep.). Leiden, The Netherlands: Department of Data Theory.

De Leeuw, J. (1994). Block relaxation algorithms in statistics. In H.-H. Bock, W. Lenski & M. M. Richter (Eds.), *Information Systems and Data Analysis* (pp. 308-324). Berlin: Springer.

De Leeuw, J., & Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate Analysis - V* (pp. 501-522). Amsterdam: North Holland.

Dykstra, R. L. (1983). An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, *78*, 837-842.

Gaffke, N., & Mathar, R. (1989). A cyclic projection algorithm via duality. *Metrika*, *36*, 29-54.

Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Chichester: Wiley.

Groenen, P. J. F., Van Os, B.-J., & Meulman, J. J. (2000). Optimal scaling by length-constrained alternating nonnegative least squares: An application to distance-based principal coordinates analysis. *Psychometrika, 65*, 511-524.

Guttman, L. (1959). A structural theory for intergroup beliefs and attitude. *American Sociological Review*, *24*, 318-328.

Heiser, W. J. (1995). Convergent computation by iterative majorization: Theory and applications in multidimensional data analysis. In W. J. Krzanowski (Ed.), *Recent Advances in Descriptive Multivariate Analysis* (pp.157-189). Oxford: Oxford University Press.

Heiser, W. J. & Meulman, J. J. (1983). Constrained multidimensional scaling, including confirmation. *Applied Psychological Measurement* *7*(4), 381-404.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115-129.

Lawson, C. L., & Hanson, R. J. (1974). *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice Hall.

Lee, S.-Y. (1984). Multidimensional scaling models with inequality and equality constraints. *Communications in Statistics - Simulation and Computation*, *13*, 127-140.

Levy, S. (1983). A cross-cultural analysis of the structure and levels of atttitudes towards acts of political protest. *Social Indicators Research*, *12*, 281-309.

Mathar, R. (1990). Multidimensional scaling with constraints on the configuration. *Journal of Multivariate Analysis*, *33*, 151-156.

Meulman, J. J., Heiser, W. J., & SPSS. (1999). *SPSS categories 10.0.* Chicago: SPSS.

Skarabis, H. (1978). Multidimensional scaling with prescribed structures. In H. Skarabis & P. Sint (Eds.), *Compstat lectures 1*. Wien, Austria: Physika Verlag.

Von Neumann, J. (1950). *Functional Operators* (Vol. 2). Princeton, NJ: Princeton University Press.

Winsberg, S., & De Soete, G. (1997). Multidimensional scaling with constrained dimensions: CONSCAL. *British Journal of Mathematical and Statistical Psychology*, *50*, 55-72.